A Boolean Algebra of Receiver Operating Characteristic Curves

Mark E. Oxley and Steven N. Thorsen
Department of Mathematics and Statistics
Air Force Institute of Technology
Graduate School of Engineering and Management
Wright-Patterson AFB, OH, 45433-7765
moxley@afit.edu, sthorsen@afit.edu

Christine M. Schubert
Department of Biostatistics
School of Medicine
Virginia Commonwealth University
Richmond, VA, 23298-0032
cmschubert@vcu.edu

***Abstract*—A reasonable starting place for developing decision fusion rules of families of classification systems is using the logical AND and OR rules. These two rules, along with the unary rule NOT, can lead to a Boolean algebra when a number of properties are shown to exist. This paper examines how these rules for classification system families comprise a Boolean algebra of systems. This Boolean algebra of families is then shown under assumptions of independence to be isomorphic to a Boolean Algebra of Receiver Operating Characteristic (ROC) curves. These decision fusion rules produce ROC curves which become the bounds by which to test non-boolean, possibly non-decision fusion rules for performance increases. We give an example to demonstrate the usefulness of this Boolean Algebra of ROC curves.**

Keywords: fusion rules, boolean algebra, receiver operating characteristic (ROC) curves, fusor, information fusion. optimization.

I. Introduction and Problem Statement

Given a finite number of families of classification systems (with 2-label output), how do we find the best possible decision fusion rule (also called label fusion rule)? In the realm of deterministic rules, Boolean rules comprise the "whole show"(almost) for decision fusion. It has been shown that under the assumption of independence of classification systems, the Boolean AND operation on families of classification systems induces another Boolean AND operation on the receiver operating characteristic (ROC) curves of each corresponding family [1]. In this paper we show this extends to the Boolean OR operation and the unary NOT operation. With these three operations defined, we develop a Boolean Algebra of ROC curves which corresponds to the Boolean algebra of a finite number of families of classification systems. This Boolean Algebra of families of classification systems is the easiest fused systems a fusion engineer can design, build, test, and evaluate. With this Boolean Algebra of ROC curves, one does not have to physically build or test the systems in order to determine its performance, and thus, determine which design is optimal.

We develop this paper first by defining families of classification systems in Section II, and review how information fusion occurs along the nodes of these families. Section III is devoted to the discussion on the fusion rules used. Section IV briefly reviews Boolean Algebras that is used in Section V, the important results, when we prove that a Boolean algebra of families of classification systems is isomorphic to a Boolean algebra of ROC curves. We give an example in Section VI, and conclude with Section VII.

Several authors have considered the Boolean Algebra of systems generated by ANDing and ORing the original systems, see [2], [3], and [4], to name a few.

II. Families of Classification Systems

The classification system can be defined mathematically, which allows the fused system to be written in terms of the individual systems. Let $\mathcal{E}$ be a population set of outcomes. Let $\mathfrak{E}$ be a σ-algebra of subsets of $\mathcal{E}$, then $(\mathcal{E}, \mathfrak{E})$ is a measurable space [5]. Let P be a probability measure defined on $\mathfrak{E}$, then $(\mathcal{E}, \mathfrak{E}, P)$ is a probability measure space. Let $\mathbf{s}$ be a sensor that produces data as its output, i.e., $\mathbf{s}$ is a mapping of outcomes from the population set $\mathcal{E}$ to a datum. Let $\mathcal{D}$ denote the data set. Then we write $\mathbf{s} : \mathcal{E} \to \mathcal{D}$ or its diagram $\mathcal{E} \xrightarrow{\mathbf{s}} \mathcal{D}$. Examples of datum from this data set may take on many forms such as infrared imagery, radar signals, data streams, or video. This data may be too difficult to classify using its current form, so a mapping $\mathbf{p}$ defined on $\mathcal{D}$ is used to produce an element x, called a feature. Typically, this element x is a vector of real numbers, though it need not be. Let the mapping $\mathbf{p}$ represent a processor that takes a datum from $\mathcal{D}$ and produces a feature, i.e., $\mathcal{D} \xrightarrow{\mathbf{p}} \mathcal{F}$. Since x might be a vector of real numbers, then $\mathcal{F} \subset \mathbb{R}^N$ for some positive integer N. Let Θ be a threshold set (or a set of parameters); maybe, $\Theta = [0, 1]$ or $\Theta = \mathbb{R} = (-\infty, \infty)$. For each $\theta \in \Theta$ let $\mathbf{a}_\theta$ be a classifier mapping $\mathcal{F}$ into a label set $\mathcal{L}$. That is, $\mathbf{a}_\theta : \mathcal{F} \to \mathcal{L}$ or $\mathcal{F} \xrightarrow{\mathbf{a}_\theta} \mathcal{L}$ for each $\theta \in \Theta$. For a two-class problem, examples of a label set could be $\mathcal{L} = \{$true, false$\}$, $\mathcal{L} = \{$T,F$\}, \mathcal{L} = \{0, 1\}$ or even $\mathcal{L} = \{$target,non-target$\}$. In this paper, we use $\mathcal{L} = \{t, n\}$ where t = "target"and n = "non-target". The graphical representation of these mappings is given by the following diagram.

$$\mathcal{E} \xrightarrow{\mathbf{s}} \mathcal{D} \xrightarrow{\mathbf{p}} \mathcal{F} \xrightarrow{\mathbf{a}_\theta} \mathcal{L} \ .$$

Define the system $\mathbf{A}_\theta$ to be the composition of these

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **JUL 2007**	2. REPORT TYPE	3. DATES COVERED **00-00-2007 to 00-00-2007**

4. TITLE AND SUBTITLE **A Boolean Algebra of Receiver Operating Characteristic Curves**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Air Force Institute of Technology,Department of Mathematics and Statistics,Graduate School of Engineering and Management,Wright-Patterson AFB,OH,45433-7765**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
10th International Conference on Information Fusion, 9-12 July 2007, Quebec, Canada.

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**8**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

mappings for each $\theta \in \Theta$. That is, for each $\theta \in \Theta$, $\mathbf{A}_\theta = \mathbf{a}_\theta \circ \mathbf{p} \circ \mathbf{s}$. Graphically, the diagram for the system is written as

$$\mathcal{E} \xrightarrow{\mathbf{A}_\theta} \mathcal{L}$$

for each $\theta \in \Theta$.

A. Two Classification Systems

There are many ways in which to express two (or more) classification systems. In this paper, however, the multiple classification system must be developed using two main premises. First, the systems to be combined are fused together using label fusion, that is, once each system has produced a label for a specific outcome from the event set, these labels are combined together to generate one overall label for that outcome. The creation of this overall label from the underlying classification systems defines how the systems are fused together, via the labels. Second, the label set for all systems considered, including each individual system and the fused classification system, contains two values or two classes. Examples of possible members of this label set were given previously, but the label set considered here is $\mathcal{L} = \{t, n\}$ where t = "target"and n = "non-target". Using the premises of label fusion and a two-class label system, representations for a two classification system are developed.

Consider the case when two sensors, $\mathbf{s}_1$ and $\mathbf{s}_2$, observe outcomes occurring in the same population set $\mathcal{E}$. Assume they produce datum in data sets $\mathcal{D}_1$ and $\mathcal{D}_2$, respectfully. That is, $\mathbf{s}_1 : \mathcal{E} \to \mathcal{D}_1$ and $\mathbf{s}_2 : \mathcal{E} \to \mathcal{D}_2$. Further, assume sensors $\mathbf{s}_1$ and $\mathbf{s}_2$ each have a processor, $\mathbf{p}_1$ and $\mathbf{p}_2$, respectively, which maps datum in the respective data sets, $\mathcal{D}_1$ and $\mathcal{D}_2$, to features in the feature sets $\mathcal{F}_1$ and $\mathcal{F}_2$. In particular, assume $\mathbf{p}_1 : \mathcal{D}_1 \to \mathcal{F}_1$ and $\mathbf{p}_2 : \mathcal{D}_2 \to \mathcal{F}_2$. Suppose that the family of classifiers for $\mathbf{p}_1$ and $\mathbf{s}_1$ is given by $\{\mathbf{a}_\theta : \theta \in \Theta\}$ and that the family of classifiers for $\mathbf{p}_2$ and $\mathbf{s}_2$ is given by another family, $\{\mathbf{b}_\phi : \phi \in \Phi\}$. Let $\mathbf{a}_\theta : \mathcal{F}_1 \to \mathcal{L}_1$ for each $\theta \in \Theta$ and $\mathbf{b}_\phi : \mathcal{F}_2 \to \mathcal{L}_2$ for each $\phi \in \Phi$. Then the labels that are produced from each of the classification systems are fused together to create an overall label for the outcome of interest. The composition of these mappings yield systems represented by the following diagram.

$$\begin{tikzcd}
 & \mathcal{D}_1 \arrow[r, "\mathbf{p}_1"] & \mathcal{F}_1 \arrow[r, "\mathbf{a}_\theta"] & \mathcal{L}_1 \\
\mathcal{E} \arrow[ur, "\mathbf{s}_1"] \arrow[dr, "\mathbf{s}_2"'] & & & \\
 & \mathcal{D}_2 \arrow[r, "\mathbf{p}_2"] & \mathcal{F}_2 \arrow[r, "\mathbf{b}_\phi"] & \mathcal{L}_2
\end{tikzcd}$$

For these two classification systems the compositions yield the systems $\mathbf{A}_\theta = \mathbf{a}_\theta \circ \mathbf{p}_1 \circ \mathbf{s}_1$ for each $\theta \in \Theta$ and $\mathbf{B}_\phi = \mathbf{b}_\phi \circ \mathbf{p}_2 \circ \mathbf{s}_2$ for each $\phi \in \Phi$. Thus, the individual diagrams are

$$\mathcal{E} \xrightarrow{\mathbf{A}_\theta} \mathcal{L}_1$$

$$\mathcal{E} \xrightarrow{\mathbf{B}_\phi} \mathcal{L}_2$$

and the two families of classification systems will be denoted by $\mathbb{A} \equiv \{\mathbf{A}_\theta : \theta \in \Theta\}$ and $\mathbb{B} \equiv \{\mathbf{B}_\phi : \phi \in \Phi\}$.

The two classification systems developed above map outcomes from the population set into different data, feature, and label sets, which are then used to fuse the classification systems together. There are, however, other ways to label outcomes from the event set. In this paper, classification systems can map outcomes into either the same or different data sets or the same or different feature sets. The sets which must remain the same for the mathematical development contained herein are the event set $\mathcal{E}$ and the two-class label set $\mathcal{L}$. Therefore, the classification systems must be acting from the same event set, map into either the same or different data and feature sets and eventually map into the same label set. That is,

$$\begin{tikzcd}
 & \mathcal{L} \\
\mathcal{E} \arrow[ur, "\mathbf{A}_\theta"] \arrow[dr, "\mathbf{B}_\phi"'] & \\
 & \mathcal{L}
\end{tikzcd}$$

B. ROC Curves

Each mapping in the classification system, as well as the composition of mappings, has a *pre-image*. Let $\mathbf{f}$ be a function mapping set $\mathcal{X}$ into set $\mathcal{Y}$, so $\mathbf{f} : \mathcal{X} \to \mathcal{Y}$. Given a subset $Y \subset \mathcal{Y}$ we define the *pre-image* of $\mathbf{f}$ to be the subset in $\mathcal{X}$ by

$$\mathbf{f}^\natural(Y) = \{x \in \mathcal{X} : \mathbf{f}(x) \in Y\}.$$

The pre-image is sometimes called the *inverse image*, although the mapping $\mathbf{f}$ need not be invertible, yet the superscript -1 is used. Because this construction creates a *natural* mapping from subsets of $\mathcal{Y}$ into subsets of $\mathcal{X}$, the natural symbol $\natural$ will be used instead of -1. Therefore, we write $\mathbf{f}^\natural(Y) = X$. If we consider the entire classification system as a composition of mappings, then we can write the pre-image of a specific label $\ell \in \mathcal{L}$ produced by the classification system $\mathbf{A}_\theta$. Let $\mathcal{L}_\ell = \{\ell\}$ so that then $\mathbf{A}_\theta^\natural(\mathcal{L}_\ell) = \{e \in \mathcal{E} : \mathbf{A}_\theta(e) \in \mathcal{L}_\ell\}$. The use of pre-images allows us to take the resulting labels and express these in terms of the underlying probabilities. This is demonstrated in the development of the ROC curve.

Assume the label set is $\mathcal{L} = \{t, n\}$ where t and n may be real values or symbols and the label t represents a "target"and the label n represents a "non-target". Define $\mathcal{L}_t = \{t\}$ and $\mathcal{L}_n = \{n\}$. We assume the event set $\mathcal{E}$ can be partitioned into a target event set containing all target outcomes and a non-target event set containing non-target outcomes. Denote the true target event set as $\mathcal{E}_t$ and the true non-target event set as $\mathcal{E}_n$. Thus, $\mathcal{E} = \mathcal{E}_t \cup \mathcal{E}_n$ and $\mathcal{E}_t \cap \mathcal{E}_n = \emptyset$.

In order to quantify how well the classification system $\mathbf{A}_\theta$ performs, we appeal to the probability measure space $(\mathcal{E}, \mathfrak{E}, P)$ to compute the following four performance quantifiers. Let $P_{TP}(\mathbf{A}_\theta)$ denote the probability of true positive classification of the classification system $\mathbf{A}_\theta$. Then $P_{TP}(\mathbf{A}_\theta)$ is the probability that the classification system $\mathbf{A}_\theta$ labels an outcome, e, as a target label, t, given that the outcome really is a target outcome from the target event set, $\mathcal{E}_t$. Mathematically, $P_{TP}(\mathbf{A}_\theta)$ is defined by the conditional probability

$$P_{TP}(\mathbf{A}_\theta) = P\{\mathbf{A}_\theta(e) = t \mid e \in \mathcal{E}_t\} = \frac{P\left(\mathbf{A}_\theta^\natural(\mathcal{L}_t) \cap \mathcal{E}_t\right)}{P(\mathcal{E}_t)}.$$

Let $P_{FP}(\mathbf{A}_\theta)$ denote the probability of false positive classification of the system $\mathbf{A}_\theta$. Then $P_{FP}(\mathbf{A}_\theta)$ is the probability that the classification system $\mathbf{A}_\theta$ labels an event outcome, e, as a target label, t, given that the outcome is really a non-target from the non-target set of the event set, $\mathcal{E}_n$. Mathematically, $P_{FP}(\mathbf{A}_\theta)$ is defined by the conditional probability

$$P_{FP}(\mathbf{A}_\theta) = P\{\mathbf{A}_\theta(e) = t \mid e \in \mathcal{E}_n\} = \frac{P\left(\mathbf{A}_\theta^\natural(\mathcal{L}_t) \cap \mathcal{E}_n\right)}{P(\mathcal{E}_n)}.$$

Let $P_{TN}(\mathbf{A}_\theta)$ denote the probability of true negative classification of the system $\mathbf{A}_\theta$. Then $P_{TN}(\mathbf{A}_\theta)$ is the probability that the classification system $\mathbf{A}_\theta$ labels an event outcome, e, as a non-target label, n, given that the outcome really is a non-target outcome from the non-target event set, $\mathcal{E}_n$. Mathematically, $P_{TN}(\mathbf{A}_\theta)$ is defined by the conditional probability

$$P_{TN}(\mathbf{A}_\theta) = P\{\mathbf{A}_\theta(e) = n \mid e \in \mathcal{E}_n\} = \frac{P\left(\mathbf{A}_\theta^\natural(\mathcal{L}_n) \cap \mathcal{E}_n\right)}{P(\mathcal{E}_n)}.$$

Let $P_{FN}(\mathbf{A}_\theta)$ denote the probability of false negative classification by the system $\mathbf{A}_\theta$. Then $P_{FN}(\mathbf{A}_\theta)$ is the probability that the classification system $\mathbf{A}_\theta$ labels an event outcome, e, as a non-target label, n, given that the outcome is really a target outcome from the target event set, $\mathcal{E}_t$. Mathematically, $P_{FN}(\mathbf{A}_\theta)$ is defined by the conditional probability

$$P_{FN}(\mathbf{A}_\theta) = P\{\mathbf{A}_\theta(e) = n \mid e \in \mathcal{E}_t\} = \frac{P\left(\mathbf{A}_\theta^\natural(\mathcal{L}_n) \cap \mathcal{E}_t\right)}{P(\mathcal{E}_t)}.$$

Note that each of these four probabilities are dependent on the threshold value, θ. A single value for each of these probabilities is computed for each value of θ. As the value of θ changes, so do the values of $P_{FP}(\mathbf{A}_\theta)$, $P_{TP}(\mathbf{A}_\theta)$, $P_{TN}(\mathbf{A}_\theta)$ and $P_{FN}(\mathbf{A}_\theta)$. Define Θ as a set of possible thresholds and for each $\theta \in \Theta$, and the set of triples

$$\tau_{\mathbb{A}} = \{(\theta, P_{FP}(\mathbf{A}_\theta), P_{TP}(\mathbf{A}_\theta)) : \theta \in \Theta\}$$

to be the *trajectory* of $\mathbb{A}$. We can project this trajectory onto the second and third component to yield the set

$$f_{\mathbb{A}} = \{(P_{FP}(\mathbf{A}_\theta), P_{TP}(\mathbf{A}_\theta)) : \theta \in \Theta\}.$$

If Θ is homeomorphic to the real numbers $\mathbb{R}$, then the trajectory $\tau_{\mathbb{A}}$ will be a curve in $\mathbb{R}^3$ and the projection $f_{\mathbb{A}}$ will be a curve in $\mathbb{R}^2$ (more specific, a curve in the unit square $[0,1] \times [0,1]$). Formally, this curve is called the *ROC curve* for the system family $\mathbb{A}$. An example of this projection is given in Figure 1. For the case when Θ is discrete, the ROC "curve"is a set of discrete points.

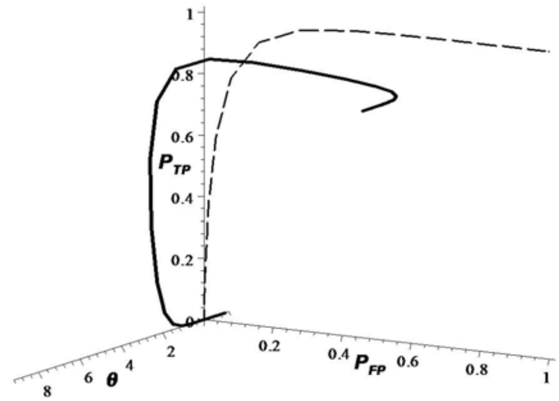

Figure 1. A ROC trajectory (solid) and its projection the ROC curve (dashed).

If Θ is a multi-dimensional set then this analysis will not yield a single curve in the P_{FP}-P_{TP} plane. Instead, a collection of curves is created. Therefore, we choose the upper frontier to be the ROC curve as representative of the classifier performance.

Definition 1: (ROC function, ROC curve) Let $\mathbb{A} = \{\mathbf{A}_\theta : \theta \in \Theta\}$ be a family of classification systems defined on the probability space $(\mathcal{E}, \mathfrak{E}, P)$ mapping to the label set $\mathcal{L} = \{t, n\}$ with parameter set Θ. For each $p \in [0,1]$, define the set

$$\Theta_p \equiv \{P_{TP}(\mathbf{A}_\theta) : \theta \in \Theta \text{ and } P_{FP}(\mathbf{A}_\theta) \leq p\}.$$

For $p \in [0,1]$, if Θ_p is nonempty then define

$$f_{\mathbb{A}}(p) = \max\{P_{TP}(\mathbf{A}_\theta) : \theta \in \Theta \text{ and } P_{FP}(\mathbf{A}_\theta) \leq p\}. \quad (1)$$

If Θ_p is empty then $f_{\mathbb{A}}(p)$ is not defined. The function $f_{\mathbb{A}}$ is called the *ROC function*. The graph of $f_{\mathbb{A}}$ is called the *ROC curve*.

In practice, the set Θ_p may be empty for certain values of p. We avoid the discussion of this case and assume that the ROC function is defined for all $p \in [0,1]$. We make this clear by defining a *total ROC function*.

Definition 2: (Total ROC function, Total ROC curve) We say a ROC curve is *total* if its ROC function is defined for all $p \in [0,1]$, that is, the ROC function is a total function.

A property of a total ROC curve are given in the following theorem.

Theorem 1: Let $\mathbb{A} = \{\mathbf{A}_\theta : \theta \in \Theta\}$ be a family of classification systems. Then $f_{\mathbb{A}}$ is a non-decreasing function. That is, for every $p, q \in [0,1]$ with $p \leq q$ then $f_{\mathbb{A}}(p) \leq f_{\mathbb{A}}(q)$.

Proof: Let $p, q \in [0,1]$ with $p \leq q$ then $\Theta_p \subseteq \Theta_q$ therefore,

$$f_{\mathbb{A}}(p) = \max_{\theta \in \Theta_p} P_{TP}(\mathbf{A}_\theta) \leq \max_{\theta \in \Theta_q} P_{TP}(\mathbf{A}_\theta) = f_{\mathbb{A}}(q).$$

■

For notational purposes we denote the collection of total ROC function by $\mathscr{R}$.

Definition 3: (Set of total ROC functions) Let the set of total ROC functions be denoted by

$$\mathscr{R} = \{f : [0,1] \to [0,1] \mid f \text{ is non-decreasing on } [0,1]\}.$$

Notice that we do not require continuity of the functions.

We write $f = g$ to mean the point-wise equality, that is, $f(p) = g(p)$ for all $p \in [0,1]$.

III. Fusion Rules

There are two types of fusion for classification systems. The first type allows for the families of classification systems which are to be fused to have exactly the same label set. We mean exactly the same, and not isomorphic, so that if the label set is, in fact, $\mathcal{L} = \{$*target, non-target*$\}$ for each family, then this means that the actual definition of a target label is identical for each. This allows for each family to partition the population set in the same way. This type of information fusion we call *within-fusion* [1].

The diagram for label fusion for two systems is

$$\begin{tikzcd}
 & \mathcal{D}_1 \arrow[r, "\mathbf{p}_1"] & \mathcal{F}_1 \arrow[r, "\mathbf{a}_\theta"] & \mathcal{L} \arrow[dd, dash] & \\
\mathcal{E} \arrow[ur, "\mathbf{s}_1"] \arrow[dr, "\mathbf{s}_2"'] & & & \| \arrow[r, dotted, two heads, "Rule"] & \mathcal{L} \\
 & \mathcal{D}_2 \arrow[r, "\mathbf{p}_2"] & \mathcal{F}_2 \arrow[r, "\mathbf{b}_\phi"] & \mathcal{L} &
\end{tikzcd}$$

Although many label-fusion rules exist, in this paper we focus on the Boolean OR and AND rules. These straightforward, "hard" rules will be used to develop a mathematical expression for the ROC curve of the fused classification system using only properties of the ROC curves from the individual systems. In this manner, if we know the performance of the individual systems, we can compute the performance of the fused system using these Boolean label-fusion rules without any replication in experimentation.

Let the ROC curve associated with the classification system family $\mathbb{A} = \{\mathbf{A}_\theta : \theta \in \Theta\}$ be denoted by $f_{\mathbb{A}}$ and the ROC curve associated with the classification system family $\mathbb{B} \equiv \{\mathbf{B}_\phi : \phi \in \Phi\}$ be denoted by $f_{\mathbb{B}}$. Recall that the label set $\mathcal{L} = \{t, n\}$.

1) AND Rule: The AND (conjunction) rule is a binary operation defined on $\mathcal{L}$. We denote the AND operation by the *join* symbol $\wedge$. Its definition is given in the table:

$\wedge$	t	n
t	t	n
n	n	n

The new classification system $\mathbf{A}_\theta \wedge \mathbf{B}_\phi$ is defined by the point-wise AND operation on its output, that is,

$$[\mathbf{A}_\theta \wedge \mathbf{B}_\phi](e) \equiv \mathbf{A}_\theta(e) \wedge \mathbf{B}_\phi(e) \quad \text{for all } e \in \mathcal{E}. \tag{2}$$

This produces a new classification system family $\mathbb{C}^{\text{AND}} = \{\mathbf{A}_\theta \wedge \mathbf{B}_\phi : \theta \in \Theta, \phi \in \Phi\}$. Thus, to be labeled as "target", both the label from systems $\mathbf{A}_\theta$ and $\mathbf{B}_\phi$ must be the "target " label. For brevity we write $\mathbb{A} \wedge \mathbb{B} = \mathbb{C}^{\text{AND}}$, thus, using the AND symbol $\wedge$ to represent the binary label AND operation (e.g., $t \wedge n$), the binary system AND operation (e.g., $\mathbf{A}_\theta \wedge \mathbf{B}_\phi$), and the binary family AND operation (e.g., $\mathbb{A} \wedge \mathbb{B}$).

2) OR Rule: The OR (disjunction) rule is also a binary operation defined on $\mathcal{L}$. We denote the OR operation by the *meet* symbol $\vee$. Its definition is given in the table:

$\vee$	t	n
t	t	t
n	t	n

Then the new classification system $\mathbf{A}_\theta \vee \mathbf{B}_\phi$ is defined by the point-wise OR operation

$$[\mathbf{A}_\theta \vee \mathbf{B}_\phi](e) \equiv \mathbf{A}_\theta(e) \vee \mathbf{B}_\phi(e) \quad \text{for all } e \in \mathcal{E} \tag{3}$$

and yields a new classification system family $\mathbb{C}^{\text{OR}} = \{\mathbf{A}_\theta \vee \mathbf{B}_\phi : \theta \in \Theta, \phi \in \Phi\}$. Thus, to be labeled as "target", either the label from system $\mathbf{A}_\theta$ or $\mathbf{B}_\phi$ must be the "target"label. For brevity we write $\mathbb{A} \vee \mathbb{B} = \mathbb{C}^{\text{OR}}$.

In comparison, the AND rule is more conservative than the OR rule in labeling of an object as target. If there are negative consequences in being labeled as a target, then a more conservative rule may be warranted in order to avoid excessive false positives. In the case of disease detection, however, the OR rule may be warranted in preventative screening, for instance, in order to avoid excessive false negatives and failure to diagnose a disease at a potentially earlier and treatable stage of development.

3) NOT Rule: The NOT (negation or complementation) rule is a unary operation defined on $\mathcal{L}$. We denote the NOT operation by $\neg$. Its definition is given in the table:

$\neg$	t	n
	n	t

Then the new classification system $\overset{\neg}{\mathbf{A}}_\theta$ is defined by the point-wise NOT operation

$$[\overset{\neg}{\mathbf{A}}_\theta](e) \equiv \overset{\neg}{\mathbf{A}}_\theta(e) \quad \text{for all } e \in \mathcal{E} \tag{4}$$

and yields a new classification system family $\overset{\neg}{\mathbb{A}} = \{\overset{\neg}{\mathbf{A}}_\theta : \theta \in \Theta\}$. Thus, to be labeled as "target", the label from system $\mathbf{A}_\theta$

must be the “non-target”. Clearly, the NOT rule is a unary rule and not a fusion rule,but it will be useful.

IV. Boolean Algebras of a finite collection of Families of Classification Systems

A. Boolean Algebras

The definition of a Boolean Algebra is given below [6].

Definition 4: A Boolean Algebra is an algebraic structure, denoted by $(\mathscr{A}, =, \wedge, \vee, \neg)$ where

- $\mathscr{A}$ is a nonempty set of elements;
- $=$ denotes element equality;
- $\wedge$ is a binary operation called AND or conjunction;
- $\vee$ is a binary operation called OR or disjunction;
- $\neg$ is a unary operation called NOT or negation (or complementation).

And the following axioms hold true:

1) $\mathscr{A}$ is closed w.r.t. $\wedge, \vee$ and $\neg$. For every $a, b \in \mathscr{A}$
$$a \wedge b \in \mathscr{A} \qquad a \vee b \in \mathscr{A} \qquad \overline{a} \in \mathscr{A} \ .$$
2) $\mathscr{A}$ is associative w.r.t. $\wedge$ and $\vee$. For every $a, b, c \in \mathscr{A}$
$$(a \wedge b) \wedge c = a \wedge (b \wedge c) \qquad (a \vee b) \vee c = a \vee (b \vee c) \ .$$
3) $\mathscr{A}$ is commutative w.r.t. $\wedge$ and $\vee$. For every $a, b \in \mathscr{A}$
$$a \wedge b = b \wedge a \qquad a \vee b = b \vee a \ .$$
4) $\mathscr{A}$ has unique identities w.r.t. $\wedge$ and $\vee$. There exists unique elements $l, u \in \mathscr{A}$ such that for every $a \in \mathscr{A}$
$$a \wedge u = a \qquad a \vee l = a \ .$$
5) $\mathscr{A}$ is absorptive w.r.t. $\wedge$ and $\vee$. For every $a, b \in \mathscr{A}$
$$a \wedge (a \vee b) = a \qquad a \vee (a \wedge b) = a \ .$$
6) $\mathscr{A}$ is distributive w.r.t. $\wedge$ and $\vee$. For every $a, b, c \in \mathscr{A}$
$$\begin{aligned} a \wedge (b \vee c) &= (a \wedge b) \vee (a \wedge c) \\ a \vee (b \wedge c) &= (a \vee b) \wedge (a \vee c) \end{aligned}$$
7) $\mathscr{A}$ contain complements w.r.t. $\wedge$ and $\vee$. For every $a \in \mathscr{A}$
$$a \wedge \overline{a} = l \qquad a \vee \overline{a} = u \ .$$

There are several other properties that follow from these axioms, see [6] for a larger list.

B. Boolean Algebra Generated from a finite number of Classification System Families

There are two special total classification system families of interest. The “target ”family $\mathbb{T}$ and the “non-target” family $\mathbb{N}$:

$$\begin{aligned} \mathbb{T} &= \{\mathbf{T}_\alpha : \alpha \in [0,1]\} \text{ and } \mathbf{T}_\alpha(e) = t \text{ for all } e \in \mathcal{E} \\ \mathbb{N} &= \{\mathbf{N}_\beta : \beta \in [0,1]\} \text{ and } \mathbf{N}_\beta(e) = n \text{ for all } e \in \mathcal{E} \end{aligned}$$

A result that is straightforward is the following.

Theorem 2: Let $\mathscr{C}$ denote the collection of total classification system families defined on the probability space $(\mathcal{E}, \mathfrak{E}, P)$ mapping to the label set $\mathcal{L} = \{t, n\}$ such that each system (in each family) is measurable with respect to $\mathfrak{E}$. Then $(\mathscr{C}, =, \wedge, \vee, \neg)$ is a Boolean Algebra.

Suppose one has K families of total classification system families that are distinct, denoted by
$$\mathscr{G} = \{\mathbb{A}^{(1)}, \mathbb{A}^{(2)}, \ldots, \mathbb{A}^{(K)}\}.$$
More specifically, assume each $\mathbb{A}^{(k)}$ cannot be produced from the other families by using the AND, OR and NOT operations. From $\mathscr{G}$ one can generate a Boolean algebra of total classification system families, denoted by $\mathscr{B}(\mathscr{G})$, taking all possible combinations of AND, OR, and NOT. This Boolean algebra is call a *Free Boolean Algebra* and $\mathscr{G}$ is called the *generator* and acts likes an “independent” set (in the same fashion as linearly independent sets are to vector spaces). From the axioms and identities, the cardinality of $\mathscr{B}(\mathscr{G})$ is 2^{2^K} [7].

The main assumption in this paper is that the classification systems are independent, that is, the occurrence or non-occurrence of an event classified by one system will not affect the occurrence or non-occurrence of another event classified by the other system. We derive expressions for the probability of true and false positive for the OR and AND label-fusion rules that can be simplified using the following definition.

Definition 5: (Independent Classification Systems) Let $(\mathcal{E}, \mathfrak{E}, P)$ be a probability space. Let $\mathcal{L} = \{t, n\}$ be a label set. Let $\mathbf{A}, \mathbf{B} : \mathcal{E} \to \mathcal{L}$ be two classification systems. We say that the systems $\mathbf{A}, \mathbf{B}$ are system independent if they are statistically independent as random variables. Thus, the collection of pre-image events are independent, so that,
$$P(\mathbf{A}^\natural(\{\ell\}) \cap \mathbf{B}^\natural(\{\ell\})) = P(\mathbf{A}^\natural(\{\ell\}))P(\mathbf{B}^\natural(\{\ell\}))$$
for all $\ell \in \{t, n\}$.

We apply this notion of independence to pre-images of the classification systems. Recall the systems $\mathbf{A}_\theta = \mathbf{a}_\theta \circ \mathbf{p}_1 \circ \mathbf{s}_1$ for each $\theta \in \Theta$ and $\mathbf{B}_\phi = \mathbf{b}_\phi \circ \mathbf{p}_2 \circ \mathbf{s}_2$ for each $\phi \in \Phi$. These compositions take sets of outcomes from the event set and map them to sets of labels in the label set. The pre-images of these non-target label sets ($\mathbf{A}_\theta^\natural(\mathcal{L}_n)$ and $\mathbf{B}_\phi^\natural(\mathcal{L}_n)$) trace the mappings back to corresponding sets in the sample space. Thus, if classification systems $\mathbf{A}_\theta$ and $\mathbf{B}_\phi$ are independent, their pre-images will be independent, as example,
$$P(\mathbf{A}_\theta^\natural(\{n\}) \cap \mathbf{B}_\phi^\natural(\{n\})) = P(\mathbf{A}_\theta^\natural(\{n\}))\,P(\mathbf{B}_\phi^\natural(\{n\})).$$

V. Boolean Algebra of ROC curves

A. AND Label-Fusion ROC Formula

Consider the development of the probabilities of true and false positive ($P_{TP}(\mathbf{C}_{\theta,\phi}^{\text{AND}})$ and $P_{FP}(\mathbf{C}_{\theta,\phi}^{\text{AND}})$, respectively) for the AND label-fusion rule under the assumption of independence.

Theorem 3: (AND Label-Fusion ROC Formula) Let $(\mathcal{E}, \mathfrak{E}, P)$ be a probability space and $\mathcal{L} = \{t, n\}$ be a label set. Let $\mathbb{A} = \{\mathbf{A}_\theta : \theta \in \Theta\}$ and $\mathbb{B} = \{\mathbf{B}_\phi : \phi \in \Phi\}$ be measurable and independent families of classification systems with admissible parameter sets, designed to classify the same target outcomes in $\mathcal{E}$. Let $f_{\mathbb{A}}$ and $f_{\mathbb{B}}$ denote their corresponding ROC curves. Let $\mathbb{A} \wedge \mathbb{B}$ be the resulting family of classification systems. Then the ROC curve $f_{\mathbb{A}\wedge\mathbb{B}}$ is given by

$$f_{\mathbb{A}\wedge\mathbb{B}}(r) = \max_{\substack{p,q\in[0,1] \\ pq=r}} f_{\mathbb{A}}(p) f_{\mathbb{B}}(q) \tag{5}$$

for every $r \in [0,1]$. Furthermore,

$$f_{\mathbb{A}\wedge\mathbb{A}} = f_{\mathbb{A}}.$$

Proof of this formula can be found in [1], [8] and [9] This formula motivates the definition of the transformation $\mathcal{T}_\wedge$ associated with the AND operation $\wedge$ and acts on the ROC curves. Specifically, given two ROC curves $f, g \in \mathscr{R}$, and for each $r \in [0,1]$ we define

$$[\mathcal{T}_\wedge(f,g)](r) = \begin{cases} \max\limits_{\substack{p,q\in[0,1] \\ pq=r}} f(p)g(q) & \text{for} \quad f \neq g \\ f(r) & \text{for} \quad f = g \end{cases} \tag{6}$$

This transformation can be shown to have the following properties:

1) (closure) $\mathcal{T}_\wedge$ is defined on all of $\mathscr{R} \times \mathscr{R}$, i.e. $\mathcal{D}(\mathcal{T}_\wedge) = \mathscr{R} \times \mathscr{R}$.
2) (associative) $\mathcal{T}_\wedge(f, \mathcal{T}_\wedge(g,h)) = \mathcal{T}_\wedge(\mathcal{T}_\wedge(f,g),h)$ for all $f,g,h \in \mathscr{R}$.
3) (symmetric) $\mathcal{T}_\wedge(f,g) = \mathcal{T}_\wedge(g,f)$ for all $f,g \in \mathscr{R}$.
4) (identity) $\mathcal{T}_\wedge(f,1) = f$ for all $f \in \mathscr{R}$.
5) (idempotent) $\mathcal{T}_\wedge(f,f) = f$ for all $f \in \mathscr{R}$.
6) (minimum element) $\mathcal{T}_\wedge(f,0) = 0$ for all $f \in \mathscr{R}$.

This shows that $\mathcal{T}_\wedge$ is a binary operation, and motivates the creation of a new symbol that represents it. Given $f, g \in \mathscr{R}$ we will write

$$f \sqcap g \equiv \mathcal{T}_\wedge(f,g).$$

We read $f \sqcap g$ as "f and g". We use a different symbol since $\wedge$ is the binary operation dealing with classification systems and $\sqcap$ deals with ROC functions.

B. OR Label-Fusion ROC Formula

Theorem 4: (OR Label-Fusion ROC Formula) Let $(\mathcal{E}, \mathfrak{E}, P)$ be a probability space and $\mathcal{L} = \{t, n\}$ be a label set. Let $\mathbb{A} = \{\mathbf{A}_\theta : \theta \in \Theta\}$ and $\mathbb{B} = \{\mathbf{B}_\phi : \phi \in \Phi\}$ be two measurable, independent families of classification systems with admissible parameter sets, designed to classify the same target outcomes in $\mathcal{E}$. Let $f_{\mathbb{A}}$ and $f_{\mathbb{B}}$ denote their corresponding ROC curves. Let $\mathbb{A} \vee \mathbb{B}$ be the resulting family of classification systems. Then the ROC curve $f_{\mathbb{A}\vee\mathbb{B}}$ is given by

$$f_{\mathbb{A}\vee\mathbb{B}}(r) = \max_{\substack{p,q\in[0,1] \\ p+q-pq=r}} [f_{\mathbb{A}}(p) + f_{\mathbb{B}}(q) - f_{\mathbb{A}}(p) f_{\mathbb{B}}(q)] \tag{7}$$

for every $r \in [0,1]$. Furthermore,

$$f_{\mathbb{A}\vee\mathbb{A}} = f_{\mathbb{A}}.$$

Proof of this formula can be found in [1].

There is a second transformation, $\mathcal{T}_\vee$, associated with the OR operation $\vee$ that acts on ROC functions. Specifically, given two ROC curves $f, g \in \mathscr{R}$, and for each $r \in [0,1]$ we define

$$[\mathcal{T}_\vee(f,g)](r) = \begin{cases} \max\limits_{\substack{p,q\in[0,1] \\ p+q-pq=r}} [f(p) + g(q) - f(p)g(q)], & f \neq g \\ f(r), & f = g \end{cases}$$

We list some properties of the transformation $\mathcal{T}_\vee$.

1) (closure) $\mathcal{T}_\vee$ is defined on all of $\mathscr{R} \times \mathscr{R}$, that is, $\mathcal{D}(\mathcal{T}_\vee) = \mathscr{R} \times \mathscr{R}$.
2) (associative) $\mathcal{T}_\wedge(f, \mathcal{T}_\wedge(g,h)) = \mathcal{T}_\wedge(\mathcal{T}_\wedge(f,g),h)$ for all $f,g,h \in \mathscr{R}$.
3) (symmetric) $\mathcal{T}_\vee(f,g) = \mathcal{T}_\vee(g,f)$ for all $f,g \in \mathscr{R}$.
4) (identity) $\mathcal{T}_\vee(f,0) = f$ for all $f \in \mathscr{R}$.
5) (idempotent) $\mathcal{T}_\vee(f,f) = f$ for all $f \in \mathscr{R}$.
6) (maximal element) $\mathcal{T}_\vee(f,1) = 1$ for all $f \in \mathscr{R}$.

We now have that $\mathcal{T}_\vee$ is a binary operation, and motivates the creation of a new symbol that represents this binary operation. Given $f, g \in \mathscr{R}$ we will write

$$f \sqcup g \equiv \mathcal{T}_\vee(f,g).$$

We read $f \sqcup g$ as "f or g". We use the symbol $\sqcup$ rather than $\vee$ in order to distinguish it from dealing with classification systems.

C. NOT ROC Formula

Given the family $\mathbb{A} = \{\mathbf{A}_\theta : \theta \in \Theta\}$ with ROC curve $f_{\mathbb{A}}$ what is the ROC curve for $\overset{\neg}{\mathbb{A}}$? Since, we have from Equation 4 that

$$\overset{\neg}{\mathbb{A}} = \{\overset{\neg}{\mathbf{A}}_\theta : \theta \in \Theta\}$$

and

$$\left[\overset{\neg}{\mathbf{A}_\theta}\right](e) = \overset{\neg}{\mathbf{A}_\theta(e)} \text{ for every } e \in \mathcal{E};$$

and, since $\mathcal{E} = \mathbf{A}_\theta^\natural(\mathcal{L}_t) \cup \mathbf{A}_\theta^\natural(\mathcal{L}_n)$, we have by the disjunctive properties:

$$\begin{aligned} P_{TP}(\overset{\neg}{\mathbf{A}_\theta}) &= \frac{P\left(\overset{\neg}{\mathbf{A}}{}_\theta^\natural(\mathcal{L}_t) \cap \mathcal{E}_t\right)}{P(\mathcal{E}_t)} \\ &= \frac{P\left(\mathbf{A}_\theta^\natural(\mathcal{L}_n) \cap \mathcal{E}_t\right)}{P(\mathcal{E}_t)} \\ &= P_{FN}(\mathbf{A}_\theta) = 1 - P_{TP}(\mathbf{A}_\theta) \end{aligned}$$

$$\begin{aligned} P_{FP}(\overset{\neg}{\mathbf{A}_\theta}) &= \frac{P\left(\overset{\neg}{\mathbf{A}}{}_\theta^\natural(\mathcal{L}_t) \cap \mathcal{E}_n\right)}{P(\mathcal{E}_n)} \\ &= \frac{P\left(\mathbf{A}_\theta^\natural(\mathcal{L}_n) \cap \mathcal{E}_n\right)}{P(\mathcal{E}_n)} \\ &= P_{TN}(\mathbf{A}_\theta) = 1 - P_{FP}(\mathbf{A}_\theta) \end{aligned}$$

Given $p \in [0,1]$ then

$$\begin{aligned} f_{\overline{\mathbb{A}}}(p) &= \max_{\theta\in\Theta}\left\{P_{TP}(\overline{\mathbf{A}}_\theta)\ :\ P_{FP}(\overline{\mathbf{A}}_\theta) \le p\right\} \\ &= \max_{\theta\in\Theta}\left\{1-P_{TP}(\mathbf{A}_\theta)\ :\ 1-P_{FP}(\mathbf{A}_\theta) \le p\right\} \\ &= 1-\min_{\theta\in\Theta}\left\{P_{TP}(\mathbf{A}_\theta)\ :\ P_{FP}(\mathbf{A}_\theta) \ge 1-p\right\} \\ &= 1-\max_{\theta\in\Theta}\left\{P_{TP}(\mathbf{A}_\theta)\ :\ P_{FP}(\mathbf{A}_\theta) \le 1-p\right\} \\ &= 1-f_{\mathbb{A}}(1-p). \end{aligned}$$

Observe that $f_{\overline{\mathbb{A}}}$ will be nondecreasing, hence, will satisfy the condition to be in $\mathscr{R}$.

Theorem 5: (NOT Label-Fusion ROC Formula) Let $(\mathcal{E}, \mathfrak{E}, P)$ be a probability space and $\mathcal{L} = \{t, n\}$ be a label set. Let $\mathbb{A} = \{\mathbf{A}_\theta : \theta \in \Theta\}$ be a total family of classification systems. Let $f_{\mathbb{A}}$ denote its corresponding ROC curve. Let $\overline{\mathbb{A}}=\{\overline{\mathbf{A}}_\theta\ :\ \theta \in \Theta\}$ be the resulting family of classification systems by the NOT operation. Then the ROC function $f_{\overline{\mathbb{A}}}$ is given by

$$f_{\overline{\mathbb{A}}}(p) = 1 - f_{\mathbb{A}}(1-p) \tag{8}$$

for every $p \in [0,1]$.

This motivates the operator $\mathcal{N}$ that acts on ROC curves $f \in \mathscr{R}$ defined by

$$[\mathcal{N}(f)]\,(p) = 1 - f(1-p)$$

for every $p \in [0,1]$.

The operator $\mathcal{N}$ satisfies the following properties:

1) (closure) $\mathcal{N}$ is defined on all of $\mathscr{R}$, that is, $\mathcal{D}(\mathcal{N}) = \mathscr{R}$.
2) (involution) $\mathcal{N}\left(\mathcal{N}(f)\right) = f$ for all $f \in \mathscr{R}$.
3) (identity) $\mathcal{N}(0) = 1, \mathcal{N}(1) = 0$.

We now have that $\mathcal{N}$ is an unary operation that acts like a negation, thus motivates the creation of a new NOT symbol acting on ROC functions (and consequently ROC curves). Given $f \in \mathscr{R}$ we will write

$$\overset{\neg}{f} \equiv \mathcal{N}(f).$$

We read $\overset{\neg}{f}$ as "not f".

Theorem 6: (ROC Boolean Algebra) $(\mathscr{R}, =, \sqcup, \sqcap, \neg)$ is a Boolean Algebra of ROC curves.

Our main result is the following theorem.

Theorem 7: Let $\mathscr{G} = \{\mathbb{A}^{(1)}, \mathbb{A}^{(2)}, \ldots, \mathbb{A}^{(K)}\}$ be a collection of K families of total classification systems that are mutually independent. Let $(\mathscr{B}(\mathscr{G}), =, \wedge, \vee, \neg)$ denote the Boolean Algebra of total, independent classification system families generates by $\mathscr{G}$. Let $\mathscr{F} = \{f_{\mathbb{A}^{(1)}}, f_{\mathbb{A}^{(2)}}, \ldots, f_{\mathbb{A}^{(K)}}\}$ be the collection of K ROC curves corresponding to $\mathscr{G}$. Then $(\mathscr{B}(\mathscr{F}), =, \sqcap, \sqcup, \neg)$ is a Boolean Algebra of ROC curves that is isomorphic to $(\mathscr{B}(\mathscr{G}), =, \wedge, \vee, \neg)$.

The Proof of this theorem is too long for this conference proceedings. We motivate its usefulness in the example in the next section.

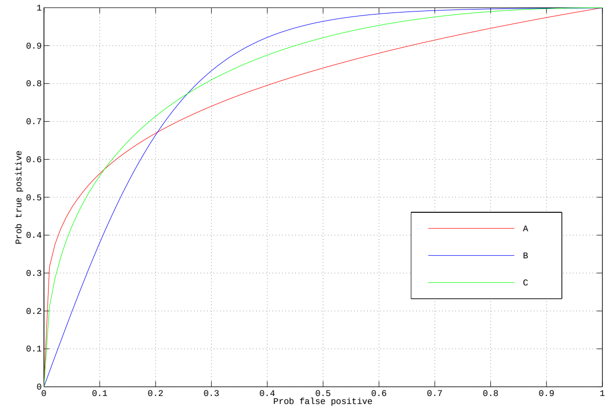

Figure 2. The ROC curves $f_{\mathbb{A}}$ (red), $f_{\mathbb{B}}$ (blue), and $f_{\mathbb{C}}$ (green).

VI. Example

Consider a problem where we know the ROC curves for the three independent classification systems families $\mathbb{A}$, $\mathbb{B}$, and $\mathbb{C}$. Assume $f_{\mathbb{A}}(p) = p^{1/4}$, $f_{\mathbb{B}}(p) = \tanh(4p)$, and $f_{\mathbb{C}}(p) = 2p^{1/2.1} - p^{2/2.1}$, see Figure (2) for their graphs. Observe that no single ROC curve completely dominates the others. The independent families generated by $\{\mathbb{A}, \mathbb{B}, \mathbb{C}\}$ of interest is given in the table below. We do not use the NOT of these families since their ROC curves will fall below the chance line implying poor performance.

single	pairs	triples
$\mathbb{A}$	$\mathbb{A}\wedge\mathbb{B}$	$(\mathbb{B}\wedge\mathbb{C})\wedge\mathbb{A}$
$\mathbb{B}$	$\mathbb{A}\wedge\mathbb{C}$	$(\mathbb{A}\wedge\mathbb{B})\vee\mathbb{A}$
$\mathbb{C}$	$\mathbb{B}\wedge\mathbb{C}$	$(\mathbb{A}\wedge\mathbb{C})\vee\mathbb{A}$
	$\mathbb{A}\vee\mathbb{B}$	$(\mathbb{B}\wedge\mathbb{C})\vee\mathbb{A}$
	$\mathbb{A}\vee\mathbb{C}$	$(\mathbb{B}\vee\mathbb{C})\vee\mathbb{A}$
	$\mathbb{B}\vee\mathbb{C}$	$(\mathbb{A}\wedge\mathbb{B})\vee\mathbb{B}$
		$(\mathbb{A}\wedge\mathbb{C})\vee\mathbb{B}$
		$(\mathbb{B}\wedge\mathbb{C})\vee\mathbb{B}$
		$(\mathbb{A}\wedge\mathbb{B})\vee\mathbb{C}$
		$(\mathbb{A}\wedge\mathbb{C})\vee\mathbb{C}$
		$(\mathbb{B}\wedge\mathbb{C})\vee\mathbb{C}$

The majority vote family is

$$\begin{aligned} \mathbb{V} &= (\mathbb{A}\wedge\mathbb{B})\vee(\mathbb{A}\wedge\mathbb{C})\vee(\mathbb{B}\wedge\mathbb{C}) \\ &= (\mathbb{A}\vee\mathbb{B})\wedge(\mathbb{A}\vee\mathbb{C})\wedge(\mathbb{B}\vee\mathbb{C}) \end{aligned}$$

We compute the ROC curves for all these families using the formulas given in Equations (5) and (7) and plot them all in Figure 3. It appears that the majority vote family dominates all the other families, but upon closer inspect we see that for small false positive values (< 0.03) the majority vote is not the best. If one chooses different ROC curves then all these curves will change, and the majority may not dominate as much. Further research will be performed to determine this dependency.

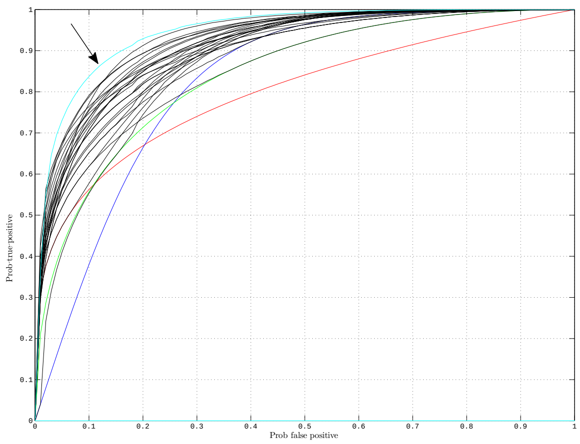

Figure 3. The ROC curves of all the families in the Boolean Algebra generated by $f_{\mathbb{A}}$ (red), $f_{\mathbb{B}}$ (blue), and $f_{\mathbb{C}}$ (green). The blacks curves are the ROC curves generated using Equations (5) and (7). The arrow points to the curve that corresponds to the majority vote family.

VII. Conclusions

We have shown that for label fusion of a finite number of classification system families, we can start with Boolean rules of AND, OR, and NOT. From this we develop a Boolean Algebra of classification system families. Under the assumption of the independence of these families, we have that this Boolean Algebra of classification system families is represented (in fact, isomorphic) by a Boolean Algebra of ROC curves. The ROC Boolean Algebra is constructed using the ROC curves of the original families. There are several possible uses for this algebra. One is that by calculating out all the elements of ROC Boolean Algebra from the original families, the cost of testing a Boolean decision rule is virtually zero. A second use is that, if we are considering a non-Boolean decision rule or a fusion rule built from the data set level or feature set level of the classification systems, by calculating the entire Boolean Algebra and taking the frontier of the resulting set of ROC curves, we can construct a bound by which to compare the performance of any other fusion rule, deterministic or randomized (see Thorsen [10]). That is, if the new fusion rule does not out perform any Boolean rule then why use it?

The applications of this procedure are manifold. A researcher is empowered to leverage legacy classification systems in ways he/she may not have thought of before, by using completely constructive testing using the ROC curves of the legacy systems.

Acknowledgments

We acknowledge the financial support for this research from the Air Force Research laboratory, Sensors Directorate, Automatic Target Recognition Branch (AFRL/SNAT) at Wright-Patterson Air Force Base, Ohio; the Air Force Office of Scientific Research, Mathematics Directorate (AFOSR/MN); and the Air Combat Command (ACC/DRCA) at Langley Air Force Base, Virginia.

The views expressed in this article are those of the authors and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the US Government.

References

[1] C. M. Schubert, “Quantifying correlation and its effects on system performance in classifier fusion,” Ph.D. dissertation, AFIT/DS/ENC/05-01 Air Force Institute of Technology, Wright-Patterson AFB OH, 2005.

[2] M. E. Liggins, C. Chong, I. Kadar, M. G. Alford, V. Vannicola, and S. Thomopoulos, “Distributed fusion architectures and algorithms for target tracking,” *Proceedings of the IEEE*, vol. 85, no. 1, pp. 95–107, 1997.

[3] M. E. Liggins, “An evaluation of CFAR effects on adaptive boolean decision fusion performance for SAR/EO change detection,” in *Signal Processing, Sensor Fusion, and Target Recognition X*, ser. Proceedings of SPIE, I. Kadar, Ed., vol. 4380, 2001, pp. 406–416.

[4] B. V. Dasarathy, *Decision Fusion*. Los Alamitos CA: IEEE Computer Society Press, 1994.

[5] S. I. Resnick, *A Probability Path*. Boston MA: Birkhauser, 2005.

[6] J. C. Abbott, *Sets, Lattices and Boolean Algebras*. Allyn and Bacon, 1969.

[7] S. Rudeanu, *Boolean Functions and Equations*. North-Holland, 1974.

[8] C. M. Schubert, M. E. Oxley, and K. W. Bauer, “A comparison of roc curves for label-fused within and across classifier systems,” in *8th International Conference on Information Fusion*, ser. Proceedings of FUSION 2005, 2005, pp. paper no. B2–3.

[9] ——, “Quantifying the performance of fused multiple classifiers,” in *Signal Processing, Sensor Fusion, and Target Recognition XIV*, ser. Proceedings of SPIE, I. Kadar, Ed., vol. 5809, 2005, pp. 390–401.

[10] S. Thorsen and M. Oxley, “A description of competing fusion systems,” *Information Fusion Journal*, vol. 7, pp. 346–360, 2006, doi:10.1016/j.inffus.2006.03.004.